

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/124747/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Corps, R E, Gambi, Chiara ORCID: <https://orcid.org/0000-0002-1568-7779> and Pickering, Martin 2020. How do listeners time response articulation when answering questions? The role of speech rate. Journal of Experimental Psychology: Learning, Memory, and Cognition 46 (4) , pp. 781-802. 10.1037/xlm0000759 file

Publishers page: <http://dx.doi.org/10.1037/xlm0000759>  
<<http://dx.doi.org/10.1037/xlm0000759>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# **How do listeners time response articulation when answering questions? The role of speech rate**

Ruth E. Corps<sup>1</sup>, Chiara Gambi<sup>1, 2</sup>, & Martin J. Pickering<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Edinburgh

<sup>2</sup> School of Psychology, Cardiff University

## **Author Note:**

The interested reader can find model outputs for precision analysis at: <https://osf.io/4btx3/>.

This research has been presented at a poster session at the 7<sup>th</sup> Joint Action Meeting and in a talk at the 23<sup>rd</sup> Architectures and Mechanisms for Language Processing Conference

## **Acknowledgements:**

Ruth Corps was supported by the Economic and Social Research Council [grant number ES/J500136/1]. Chiara Gambi was supported by a Leverhulme Research Project Grant [RPG-2014-523] to Martin Pickering and Hugh Rabagliati. We thank Sven Mattys for helpful suggestions regarding speech rate manipulations, Nigel Corps for recording stimuli audio, and Max Dunn for acting as second coder for both experiments.

Please address correspondence to:

Ruth Elizabeth Corps

Department of Psychology

7 George Square

University of Edinburgh

Edinburgh EH8 9JZ

United Kingdom

rcorps@exseed.ed.ac.uk

Word count: 12818 (excl references & appendix)

## Abstract

During conversation, interlocutors often produce their utterances with little overlap or gap between their turns. But what mechanism underlies this striking ability to time articulation appropriately? In two verbal *yes/no* question-answering experiments, we investigated whether listeners use the speech rate of questions to time articulation of their answers. In Experiment 1, we orthogonally manipulated the speech rate of the context (e.g., *Do you have a...*) and final word (e.g., *dog?*) of questions using time-compression, so that each component was spoken at the natural rate or twice as fast. Listeners responded earlier when the context was speeded rather than natural, suggesting they used the speaker's context rate to time answer articulation. Additionally, listeners responded earlier when the speaker's final syllable was speeded than natural, regardless of context rate, suggesting they adjusted the timing of articulation after listening to a single syllable produced at a different rate. We replicated this final word effect in Experiment 2, which also showed that our speech rate manipulation did not influence the timing of response preparation. Together, these findings suggest listeners use speech rate information to time articulation when answering questions.

**Keywords:** speech rate, time compression, prediction, turn-taking, question-answering

## Introduction

Accurately predicting when future events will occur is important for many successful interactions. People often predict their partner's timing by tracking temporal regularities in their partner's behavior, for example in their actions (e.g., Pecenka & Keller, 2011) or in their speech (e.g., Arnal & Giraud, 2012). Accurate timing predictions are likely to be particularly important in natural conversation, in which speakers' contributions are so finely coordinated that there is often little overlap or gap between their turns (mode of around 200 ms; Stivers et al., 2009). But are the mechanisms underlying predictive timing in language production similar to those used in language comprehension? We address this question in two experiments and ask whether the speech rate of comprehended questions influences when participants initiate articulation of their answers.

There is much evidence that many of the representations used during language comprehension can influence those used during language production. For example, comprehending word primes that are semantically or associatively related to the name of a target picture affects naming times during production (e.g., Alario, Segui, & Ferrand, 2000; Schriefers, Meyer, & Levelt, 1999). Additionally, interlocutors often align their representations, such that they tend to repeat each other's choice of syntactic structure (Branigan, Pickering, & Cleland, 2000) and referring expressions (e.g., Brennan & Clark, 1996). Finally, studies investigating syntactic repetition using fMRI have shown that the same brain areas are activated in comprehension and production (e.g., Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012). Although it is unclear whether phonological representations are shared across modalities (see Gambi & Pickering, 2017), it certainly appears that representations of lexico-syntactic content (i.e., what the speaker is going to say) activated during comprehension can influence later production, and vice versa.

Other research suggests that representations of timing are also shared across production and comprehension (and perhaps across action and perception more generally; e.g., Knoblich, Butterfill, & Sebanz, 2011). For example, Jungers and Hupp (2009; Jungers, Palmer, & Speer, 2002) found that listeners were more likely to produce picture descriptions at a fast rate after hearing a prime sentence produced at a fast rate. Similar results were reported in dialogue by Schultz, O'Brien, Phillips, and McFarland (2016; see also Street, 1984), who found that interlocutors' speech rates became mutually entrained during scripted turn-taking conversations: Participants produced their turn at a faster beat rate after their interlocutor produced their own turn at the same fast beat rate.

Together, these studies suggest that speech rate can be primed, such that the rate of utterances during comprehension can influence the rate at which subsequent speech is produced. But these studies have not investigated whether the speech rate of a speaker's turn influences when listeners time the initiation of their turn. In other words, does the speech rate of comprehended utterances affect the timing with which listeners *initiate* articulation? Our experiments used a question-answering task to investigate whether the timing of when participants initiate articulation of their verbal answers is influenced by the speech rate of a speaker's question.

Speakers often vary in their speaking rates (e.g., Tauroza & Allison, 1990; Miller & Dexter, 1988; Miller, Grosjean, & Lomanto, 1984), and so listeners must take this information into account if they wish to produce their own turn at the appropriate moment, so they do not overlap or leave long gaps between utterances. Indeed, two theoretical accounts of conversational turn-taking argue that speech rate can affect the duration of inter-turn intervals. First, Wilson and Wilson (2005) claimed that listeners entrain to (or track) an interlocutor's speech rate using cyclic neural oscillators, which are pools of neurons that synchronize to an external rhythm (Large & Jones, 1999). Indeed, much evidence suggests

that neural oscillators underlie speech rate entrainment during comprehension. For example, Zion Golumbic et al. (2013; see also Ding et al., 2017) recorded electrocorticographic (ECoG) activity in the auditory cortex while listeners attended to one of two speakers. They found that oscillations in both the high and low frequency ranges tracked the rhythm of the attended speech. Further studies suggest that these oscillators are sensitive to the speaker's rate of syllable production. For example, Doelling, Arnal, Ghitza, and Poeppel (2014; see also Ghitza, 2012) found that the correspondence between oscillatory activity and the speech signal was reduced when temporal fluctuations associated with syllable rate were removed. Such correspondence was regained when these fluctuations were artificially reinstated by inserting silent gaps, so that the syllable rate of the manipulated turn was comparable to that of the natural turn.

Wilson and Wilson (2005) argued that conversational overlap is rare because interlocutors' oscillatory cycles are entrained in anti-phase, so that the listener's (i.e., the next speaker) readiness to produce a syllable is at a maximum when the current speaker's readiness is at a minimum (and vice versa; but see Cummins, 2012, for a discussion of why syllables may not underlie entrainment). In the context of turn-taking, anti-phase means that listeners will be maximally ready to produce their turn half a syllable before or after the end of the speaker's turn. If listeners do not produce a response half a syllable before the end of a turn, then they will not be able to begin speaking again until they have completed another oscillatory cycle (i.e., for the duration of another cycle). This account therefore predicts that listeners will be maximally ready to produce their response half a syllable before or after the end of the speaker's turn, meaning that inter-turn intervals should be bimodally distributed around zero.

But there is no evidence to support Wilson and Wilson's argument that interlocutors' oscillatory cycles are in anti-phase. For example, Beňuš (2009) tested the oscillator theory

using 12 dyadic conversations between people playing computer games from the Columbia Games Corpus. Turn intervals were unimodally (rather than bimodally) distributed, with a peak around 100-200 ms, which is consistent with research demonstrating that turn-intervals typically last between 0 and 200 ms in a variety of different languages (Stivers et al., 2009). Furthermore, although support for Wilson and Wilson's (2005) account can be drawn from studies demonstrating speech rate priming (e.g., Jungers & Hupp, 2009), another study has found that speech rate convergence does not influence the duration of inter-turn intervals (see Finlayson, Lickley, & Corley, 2012).

In a second theoretical account, Garrod and Pickering (2015) also proposed that speech rate affects the duration of inter-turn intervals. Specifically, they argued that listeners use the rate of the speaker's preceding syllables to predict the rate of their forthcoming syllables. This prediction then affects when listeners launch articulation, such that turn transitions should be shorter when the speaker produces their turn at a faster than a slower rate, because listeners should predict that they can launch articulation earlier. Unlike Wilson and Wilson (2005), however, this account does not make any claim about interlocutors' cycles being in anti-phase, and instead allows for many other factors to affect the duration of turn intervals. In fact, research suggests that determining when to launch articulation likely depends on a number of mechanisms, such as predicting the speaker's turn-end (e.g., De Ruiter, Mitterer, & Enfield, 2006) and reacting to phonetic cues that signal turn-finality (such as downstepping, lengthening of the final word, or a drop in pitch; e.g., Bögels & Torreira, 2015; Gravano & Hirschberg, 2011). Although turn-end prediction and turn-final cues are not the focus of our experiments, we controlled for their presence to ensure they did not differ across conditions.

Consistent with Garrod and Pickering's (2015) account, a number of studies suggest that listeners can use an interlocutor's speech rate to predict the rate of their forthcoming



syllables. In one study, Dilley and Pitt (2010; see also Pitt, Szostak, & Dilley, 2016) either expanded (by a factor of 1.9) or compressed (by a factor of 0.6) the rate of the context surrounding a co-articulated single-syllable function word (e.g., *or* in *Deena doesn't have any leisure or time*). When context rate was slowed, listeners often failed to perceive this function word (*leisure or time* was perceived as *leisure time*); when context rate was speeded, listeners often erroneously perceived an absent function word (*leisure time* was perceived as *leisure or time*). Dilley, Morrill, and Banzina (2013) reported a similar pattern of results with reduced syllables, suggesting that this effect is not limited to function words. This disappearing-syllable effect may occur because the listener has entrained to the speaker's syllable rate and predicts that future syllables will continue to be produced at the same rate. This prediction then causes the listener to adopt the interpretation that is more compatible with the predicted rate, leading to the loss or insertion of a syllable. In support of this proposal, Kösem et al. (2017) found that low frequency activity in the auditory cortex entrained to the context rate of a sentence and, crucially, was sustained *after* a rate change occurred.

Similar results have also been found over longer timescales. Using the same procedure as Dilley and Pitt (2010), Baese-Berk, Dilley, Schmidt, Morrill, and Pitt (2014; see also Morrill, Dilley, McAuley, & Pitt, 2014) manipulated both the speech rate of individual utterances (the distal rate) and the average speech rate of utterances across the whole experiment (the global rate). They found that participants were less likely to perceive a function word when the context rate of an individual utterance was slowed, thus replicating Dilley and Pitt's results. In addition, listeners were less likely to perceive a function word when global speech rate was slower, suggesting that the disappearing-word effect was also influenced by the rate of utterances across the whole experiment. Together, these results do not only confirm that speech rate affects listeners' timing predictions, but also that such

predictions are based on speech rate over different timescales: both over the course of a speaker's individual utterance, and over many utterances.

Studies demonstrating a disappearing-word effect suggest that listeners use their interlocutor's syllable rate to predict that forthcoming syllables will continue at the same rate. However, we do not know whether these predictions made during language comprehension can affect when interlocutors initiate articulation during language production, as suggested by Garrod and Pickering (2015). We investigated this issue in two experiments by presenting participants with simple questions (e.g., *Do you have a dog?*) and instructing them to answer either *yes* or *no*. In Experiment 1, we used a method similar to Pitt et al. (2016; see also Dilley & Pitt, 2010) and orthogonally manipulated the context (e.g., *Do you have a...*) and final word (always consisting of a single syllable; e.g., *dog?*) rate of questions. We used either a natural rate (normal spoken rate) or a speeded rate (compressed by a factor of 0.5, so it was twice as fast as its natural rate).

If the speech rate of utterances during comprehension affects when interlocutors initiate articulation in subsequent language production, then we expect the speech rate of the context of the speaker's question to influence participants' answer initiation times (i.e., the onset of their answer). In particular, we expect participants to use the context rate of the speaker's question to predict that the speaker will produce their final syllable (i.e., the final word) at the same rate. As a result, participants should answer later when the speaker produces the context at a normal rather than a speeded rate because they predict that the speaker will reach the end of their final syllable later, consistent with studies demonstrating prediction of timing during comprehension (e.g., Dilley & Pitt, 2010).

These results would be consistent with a *shared timing* account, in which timing representations in comprehension can affect language production. This account makes a number of predictions regarding possible effects of final word rate. Research suggests that

comprehenders make timing predictions based on speech rate over multiple timescales (e.g., Baese-Berk et al., 2014), and some theories suggest that listeners adjust their timing predictions on a syllable-by-syllable basis (e.g., Giraud & Poeppel, 2012; Peelle & Davis, 2013). If this is the case, then we may expect participants to adjust their timing predictions when they encounter a syllable produced at a different rate from the question context. As a result, participants will answer earlier when the speaker produces their final word at a faster rate than the rest of the question because participants adjust their timing prediction upon encountering this final syllable and predict that the question will end earlier than they had expected based on context alone. Conversely, participants will answer later when the speaker produces their final word at a slower rate than the rest of the question.

Alternatively, it is possible that changes in speech rate during comprehension (for example, when the speaker suddenly changes their rate of syllable production) do not immediately affect the timing of articulation initiation during language production. Previous studies showing that the speech rate of utterances during comprehension primes the rate of utterances during production have kept the rate of utterances consistent throughout turns (e.g., Jungers & Hupp, 2009). As a result, changes in rate during comprehension may not affect the timing of articulation initiation during subsequent production. If this is the case, then answer times will be influenced by the context rate of questions, but not by the rate of the final word.

Of course, it is also possible that timing representations activated during comprehension do not influence production (and vice versa; *separate timing* account). If this is the case, then we expect no effects of context or final word rate on the timing of participants' responses. This account appears unlikely, however, since previous studies demonstrating speech rate convergence (e.g., Jungers & Hupp, 2009) suggest that the speech

rate of utterances during comprehension can prime the rate of utterances during language production.

Studies assessing the timing of utterances during question-answering typically measure answer times from the offset of the speaker's final word (i.e., utterance end; e.g., Bögels, Magyari, & Levinson, 2015; Corps, Crossley, Gambi, & Pickering, 2018), as this measure is equivalent to inter-turn intervals in natural dialogue. However, we measured answer times from the onset of the speaker's final word, because answer times from final word offset are affected by response preparation: Participants could respond more closely to final word onset when this word is longer simply because they have more time for response preparation (see Magyari, De Ruiter, & Levinson, 2017). If this preparation advantage following natural words is sufficiently large, it may even mask any effect of final word rate. Instead, analyses from final word onset are unlikely to be affected by response preparation to the same extent.

Note that any effect of final word rate in Experiment 1 would be consistent with the possibility that participants make timing predictions over multiple timescales, but it could also be explained by other factors. In particular, participants may answer speeded final words earlier than natural final words simply because speeded words are recognized earlier, which in turn allows participants to start preparing a response earlier. Thus, we further tested for an effect of final word rate in Experiment 2, in which a manipulation of final word rate was crossed with a manipulation of content predictability. By making the final word predictable in half of the questions, we allowed participants to start response preparation before a rate change occurred and before they even heard the final word. If listeners make timing predictions over multiple timescales, and our final word manipulation affects the timing of articulation and not the timing of response preparation, then we expect the effect of final word rate in Experiment 2 to be comparable for the predictable and unpredictable conditions.

## Experiment 1

To determine whether speech rate in comprehension influenced answer times in production, we used a verbal *yes/no* question-answering task and manipulated the speech rate of these questions using time compression, so that the context (e.g., *Do you have a...*) and final word (made up of a single syllable; e.g., *dog?*) were either compressed by a factor of 0.5 (i.e., twice as fast as the natural spoken rate; *speeded* conditions) or presented at the spoken rate (*natural* conditions). In other words, we created four conditions in which a natural or speeded context was combined with a natural or speeded final word (natural-natural, natural-speeded, speeded-speeded, and speeded-natural conditions).

This manipulation was designed to test whether the speech rate of utterances during comprehension can affect the timing of articulation initiation during subsequent language production. The *shared timing* account predicts that participants will initiate articulation earlier when the context of a question is produced at a speeded rather than a natural rate. If the timing of articulation initiation is influenced by timing predictions made over multiple timescales, then we also expect participants to answer earlier when the final word of the question is produced at a speeded than a natural rate, regardless of context rate. If, however, changes in speech rate during comprehension do not immediately affect language production, then answer times should not be influenced by final word rate. The *separate timing* account, in contrast, predicts that timing representations in comprehension do not influence production, and so we expect answer times to be unaffected by context and final word rate.

Although neither of these accounts predict an interaction between context and final word rate, we included this interaction in our analyses to control for other factors that may affect our results. First, participants may be surprised to encounter a rate change, and so we might expect a smaller effect of final word rate when a rate change occurs and the context is

natural (natural-speeded vs. natural-natural condition) rather than speeded (speeded-natural vs. speeded-speeded condition). When a natural context is followed by a speeded final word (natural-speeded), listeners should respond earlier than when the final word is also produced at a natural rate (natural-natural condition). However, these earlier responses will be counteracted by the slowing down effect of surprise (because a speeded final word comes after a rate change). But when a speeded context is followed by a natural final word (speeded-natural condition), listeners should respond even later than would be expected based on final word rate alone, because the delay in responses that we expect as a result of surprise adds to the delay that we expect after a natural final word (i.e., compared to speeded-speeded).

Furthermore, some research suggests that final lengthening can act as a turn-yielding cue (e.g., Gravano & Hirschberg, 2011). If listeners in our experiments use final lengthening as a cue to start articulation of their response, we expect an interaction between final word rate and context rate. In particular, participants should respond earlier in the speeded-natural condition, in which the final word is lengthened in comparison to the rest of the utterance, than in the speeded-speeded condition. Final lengthening is not present in the natural-natural and natural-speeded conditions, and so we expect no difference in response times in these two conditions.

## **Method**

### *Participants*

Thirty-two native English speakers (4 males; *M*<sub>age</sub> = 19.44) at the University of Edinburgh participated in exchange for course credit or £4. Participants had no known speaking, reading, or hearing impairments. The experiment was approved by the Ethics Committee of the Department of Psychology, University of Edinburgh.

## *Materials*

Participants listened to 124 questions. All questions were recorded by a native English male speaker, who was instructed to read the utterances as though “you are asking a question and expecting a response”. Since previous research suggests that prosodic cues play a role during turn-taking (e.g., Duncan, 1972), we inspected our audio recordings for such cues both auditorily and phonetically (i.e., waveform and spectrogram) using Praat (Boersma & Weenink, 2002). All questions had falling boundary tones. Boundary tone judgements were validated by a second coder, who listened to a randomly selected 25% of the utterances, which resulted in a Cohen’s kappa of 1 (i.e., complete agreement). In addition, some research suggests that pitch downstep, which occurs when the pitch of each syllable is lower than the previous syllable (Beckman & Pierrehumbert, 1986), can act as a turn-yielding cue (Cutler & Pearson, 1986), meaning that we may expect listeners to respond earlier when downstepping is present. Although two independent raters could not agree on downstep (Cohen’s kappa < .75) judgments for the stimuli, this disagreement should not pose a problem for interpretation of the results, given that the manipulation is within-items and time compression does not alter the pitch of utterances. These downstep ratings and boundary tone judgments will become relevant in Experiment 2, however, where our manipulation is between items.

We manipulated the speech rate of these questions using a time-compression factor of 0.5. Stimuli were time compressed using the Pitch-Synchronous Overlap and Add (PSOLA) algorithm in Praat (Moulines & Charpentier, 1990). This method altered utterance speech rate (so it was produced twice as fast, i.e., speeded utterances; see Table 1) but left the speech stream unaltered in the frequency-domain (preserving e.g., pitch and segmental information). Both the natural and speeded utterances were divided into a context and a final word (which included any pause prior to the onset of the final word) to create two versions of each

(speeded context, natural context; speeded final word, natural final word). Context and final word regions were then recombined to create four stimuli conditions: (i) natural-natural,

Context	Final		Total Duration	Context Duration	Final Word Duration
Natural	Natural	<i>M</i>	1838	1341	497
		<i>SD</i>	416	418	96
	Speeded	<i>M</i>	1591	1341	250
		<i>SD</i>	414	418	48
Speeded	Natural	<i>M</i>	1166	669	497
		<i>SD</i>	217	208	96
	Speeded	<i>M</i>	910	669	250
		<i>SD</i>	218	208	48

where both the context and the final word were presented at the spoken rate; (ii) natural-speeded, where the context was presented at the spoken rate, but the final word was compressed; (iii) speeded-speeded, where both the context and the final word were compressed; and (iv) speeded-natural, where the context was compressed, but the final word was presented at the spoken rate.

Table 1: The means (*M*) and standard deviations (*SDs*) of the total duration, context duration, and final word duration (ms) for the four stimuli conditions.

To ensure our time compression manipulation did not make the sentences unintelligible (given that participants were expected to comprehend the questions before



answering), we assessed intelligibility using a pre-test, in which 28 further participants (6 males;  $M_{age} = 19.61$ ) listened to the questions and typed exactly what they heard the speaker say. We calculated the average intelligibility of each utterance by comparing the number of words in the question to the number of words participants correctly identified. Any obvious spelling mistakes or typing errors (i.e., from keys around the target letter or missing letters) were scored as correct, but morphological mismatches were not (e.g., *younger* would be scored incorrect if the target was *young*; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Loebach, Pisoni, & Svirsky, 2010). Importantly, intelligibility was high (>98%) in all four conditions (mean of 99.6% in the natural-natural condition, 99.2% in the natural-speeded condition, 99.1% in the speeded-natural condition, and 98.9% in the speeded-speeded condition), and so questions were easily comprehended. However, intelligibility was lower in the speeded than the natural context conditions (Context Rate:  $F(1, 492) = 6.04, p = .01$ ; Final Word Rate:  $F(1, 492) = 3.39, p = .07$ ; Interaction:  $F(1, 492) = 0.11, p = .74$ ). If intelligibility influences answer times, then we expect participants to answer later in the speeded context conditions (i.e., the opposite of what the shared timing account suggests).

Previous work indicates that listeners respond earlier when content is predictable than when it is unpredictable (e.g., Corps et al., 2018; see also Bögels et al., 2015). To limit between-items variability, we selected only questions that were unpredictable in content. We selected these stimuli using a cloze pre-test (Taylor, 1953), in which 21 further participants from the same population (3 males,  $M_{age} = 21.43$ ) were presented with questions missing their final word and were instructed to “complete each fragment with the word or words that you think are most likely to follow the preceding context of the question.” We assessed the constraint of question fragments (i.e., the extent to which fragments predicted a particular continuation) using Shannon entropy (i.e.,  $-\sum p_i \log_2(p_i)$ , where  $p_i$  is the proportion of times

each completion occurred for a given fragment; Shannon, 1948), which is low (a minimum of 0) when completions are the same across participants and the question predicts a particular continuation (i.e., content is predictable), and high (a maximum of 4.39 when each of the 21 participants in the pre-test provided a different completion) when completions are different. Content entropy was high (see Table 2), indicating that questions did not predict a particular continuation. We also assessed the predictability of the final word of questions using cloze probability (Taylor, 1953) to calculate the percentage of participants who provided a particular continuation for a given question fragment. All final words we selected had low cloze probability (see Table 2).

Note that the data from this pre-test also allowed us to check that question fragments did not differ in length predictability (the number of words that participants would expect to complete them). We calculated length predictability using entropy (using the same formula for content entropy, but  $p_i$  is the proportion of times each completion length occurs for a given fragment), which was low for all fragments (see Table 2). Furthermore, all questions were completed with a single word by at least 70% of participants and so one word completions were predictable for all our stimuli (see Magyari & De Ruiter, 2012). All completions consisted of a single monosyllabic word, to ensure that the final word of all stimuli provided participants with the same amount of information (i.e., a single syllable) about a change of rate.

Table 2: The means ( $M$ ) and standard deviations ( $SD$ ) of our measures of content predictability, length predictability, difficulty, and plausibility for the 124 stimuli

	$M$	$SD$
Completion Length Entropy <sup>a</sup>	0.63	0.39
Completion Length Cloze <sup>b</sup>	86%	9%

Completion Content Cloze <sup>c</sup>	6%	3%
Content Entropy <sup>d</sup>	3.28	0.63
Question Difficulty <sup>e</sup>	6.16	0.05
Question Plausibility <sup>e</sup>	5.82	0.08

---

<sup>a</sup> Entropy of the length (in number of words) of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion length.

<sup>b</sup> Percentage of participants who provided the word length of the selected completion used in the main experiment (always a single word) as a continuation in the cloze task.

<sup>c</sup> Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

<sup>d</sup> Entropy of the content of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.

<sup>e</sup> Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, and 7 indicated that the question was very plausible/easy to answer.

Finally, we measured the difficulty and plausibility (see Table 2) of all questions using ratings obtained from a second pre-test, in which 12 further participants (6 males;  $M_{age} = 29.92$ ) rated (i) how difficult they would find it to answer the question if asked, and (ii) whether the question made sense. Both ratings were made on a scale of 1 (very implausible/difficult) to 7 (very plausible/easy). The mean ratings of 6.16 for difficulty and 5.82 for plausibility indicated that the questions were judged to be fairly easy and plausible.

### *Design*

Both context rate (speeded vs. natural) and final rate (speeded vs. natural) were varied within participants and items, and so there were four versions of each stimulus. We created four experimental lists (each containing 124 questions) using a Latin Square procedure, so that all participants saw one version of each item and 31 items from each condition.

### *Procedure*

Stimulus presentation and data recording were controlled using E-Prime (version 2.0). A fixation cross (+) appeared 500 ms before question onset, and the screen turned red as audio playback began. Participants pressed a button on the response box to start audio playback of the question, and were told to: “Answer either *yes* or *no* as quickly as possible. Do not wait until the speaker has finished the question and has stopped speaking. Instead you should answer as soon as you expect the speaker to finish.” Thus, participants were encouraged to respond as quickly as possible. Participants responded using the microphone provided, and playback stopped as soon as a voicekey response was recorded.

At the start of the experiment, participants completed twelve practice trials (three from each of the four conditions) to familiarize themselves with the experimental procedure. The 124 stimuli were individually randomized for each participant, and participants were given the opportunity to take a break after every 31 items.

### **Data Analysis**

Of the 3968 answers, 175 (4.41%) were discarded because the audio recording was unclear and so the answer could not be categorized as either *yes* or *no*. We removed a further three (0.08%) answer times greater than 10000 ms because they were clear outliers. We then replaced any responses falling 2.5 standard deviations above (80; 2.02%) or below (27; 0.68%) the by-participant mean answer time with the respective cut-off value.

We first calculated answer times from final word offset (i.e., question offset), as this measure is equivalent to inter-turn intervals in natural dialogue. However, our primary dependent variable was response time measured from final word onset, which was derived by adding final word duration to the participants' response time from question offset. We assume that participants prepared their own response after the onset of the final word of the speaker's question, because all of our questions were unpredictable in content and participants could determine what to respond only after the speaker began producing the critical final word. But since participants prepared their response after the onset of the final word, then it also means that a larger amount of response preparation could be carried out before the offset of the final word when this final word was longer. As a result, we would expect participants to respond closer to final word offset when this word is longer (see Magyari et al., 2017, for a similar argument). Indeed, there was a negative correlation between final word duration and answer time from question offset ( $r = -0.21, p < .001$ ), such that questions with longer final words tended to elicit earlier responses than those with shorter final words.

As a consequence, analyses from final word offset may not be informative about the use of speech rate information for timing articulation: Participants may respond closer to the offset of natural than speeded final words simply because natural words are longer, which gives them more time to prepare an answer. If this preparation advantage following natural words is sufficiently large, it may even mask any effect of final word rate. Instead, analyses from final word onset are unlikely to be affected by response preparation to the same extent, and thus provide a better index of the effect of speech rate on timing articulation initiation.

To check our assumptions, as well as testing our hypotheses, we evaluated the effects of context rate and final word rate on answer times from both final word onset and offset with linear mixed effects models (LMM; Baayen, Davidson, & Bates, 2008) using the *lmer*

function of the *lme4* package (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in RStudio (version 0.99.896) with a Gaussian link function. In all instances, we fitted models using the maximal random effects structure justified by our design (Barr, Levy, Scheepers, & Tily, 2013), but correlations among random effects were fixed to zero to aid model convergence (Matuschek, Kliegel, Basisht, Baayen, & Bates, 2017). We fitted the full model, in which answer times (from either final word onset or offset) were predicted by Context Rate (reference level: natural vs. speeded), Final Word Rate (reference level: natural vs. speeded), and their interaction.

To account for other factors that may affect answer times, we also included interactions between Context Rate, Final Word Rate, and two control predictors. Research suggests that participants tend to answer *yes* more quickly than *no* (e.g., Strömbergsson, Hjalmarsson, Edlund, & House, 2013), and so we included Answer (reference level: no vs. yes). In addition, we included Answer Agreement, which is the absolute difference between the percentage of participants who answered *yes* and the percentage who answer *no* (i.e., with 100 occurring if all answered *yes* or all answered *no*, and 0 occurring if half answered each way), as participants answer quicker when Answer Agreement is higher (Corps et al., 2018). Note that we included interactions between Answer and Answer Agreement, but we do not report two-way interactions between these predictors or four-way interactions between these predictors and Context Rate and Final Word Rate (full models are reported in the Appendix) for ease of reading. In other words, we report only two- and three-way interactions between Context Rate, Final Word Rate, and either Answer or Answer Agreement.

Finally, some studies have found a negative relationship between duration of the whole turn (not just the final word) and response times (e.g., Corps et al., 2018; Corps, Pickering, & Gambi, 2018; De Ruiter et al., 2006), and so we also included Question Duration as a main effect. We did not include interactions between Question Duration and

Context Rate and/or Final Word Rate, since this predictor likely captures some of the same variance as Context Rate and Final Word Rate: Contexts and final words that are speeded will have a shorter duration than those that are natural.

In all instances, control predictors were included only in the fixed effects structure, and not as random effects, to aid model convergence. Thus, our analyses included random effects for Context Rate, Final Word Rate, and their interaction for both participants and items. All predictors were contrast coded (-0.5, 0.5; where relevant) and centered before being added to the model. We assume that a  $t$  value greater than 1.96 indicates significance at the 0.05 alpha level (Baayen et al., 2008), and we report coefficient estimates ( $b$ ), standard errors ( $SE$ ), and  $t$  values for each predictor of interest. We computed Bayes Factors (reported as BF in the Results section) for experimental predictors by fitting Bayesian mixed effects models as implemented in the *brms* package (version 2.6.0). We report models fitted using a normal distribution, with 10000 iterations, and Cauchy priors with a mean of zero and a standard deviation of one (see Rouder, Morey, Speckman, & Province, 2012). Following Dienes (2014), we interpret a Bayes Factor (i) greater than 3 as strong evidence for the alternative hypothesis over the null, (ii) less than 0.33 as strong evidence for the null hypothesis over the alternative, and (iii) between 0.33 and 3 as weak evidence.

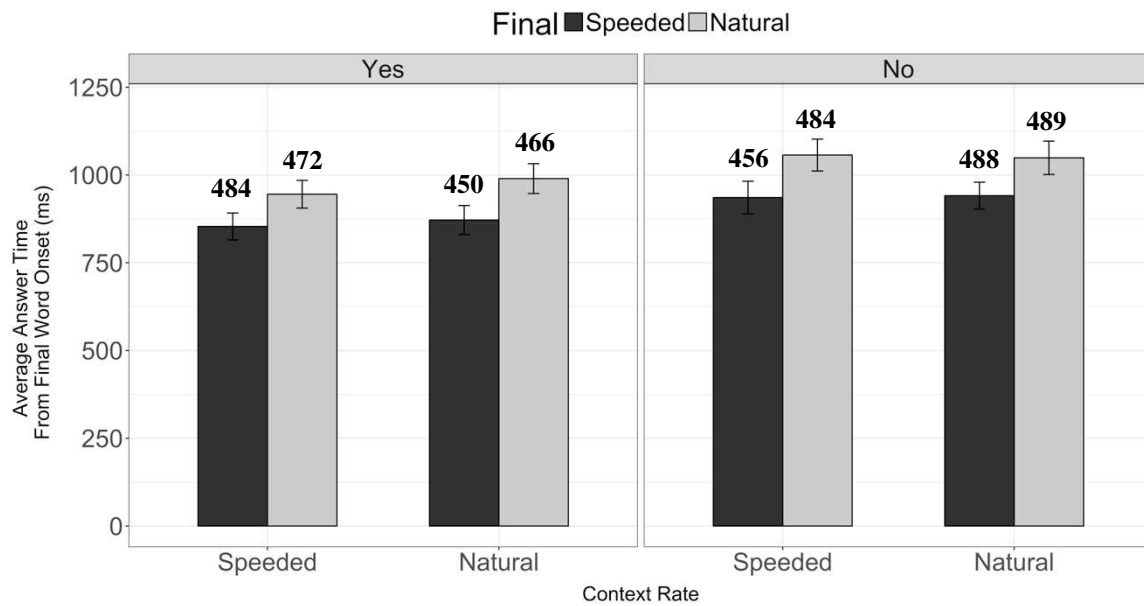
## Results

### *Analysis from final word onset: Speech rate and articulation initiation*

Consistent with the shared timing account, we found an effect of Context Rate: Participants answered earlier after a speeded than a natural context ( $b = -44.35$ ,  $SE = 190.07$ ,  $t = -2.33$ ,  $BF = 295$ ; mean answer times for speeded = 947 ms vs. natural = 966 ms; see Figure 1). But in addition, participants answered earlier when the final word was speeded than when it was natural ( $b = -122.95$ ,  $SE = 13.74$ ,  $t = -8.95$ ,  $BF = 95721635$ ; mean answer times for speeded = 899 ms vs. natural = 1012 ms).

Figure 1

Average answer times (ms) from final word onset in the four conditions for yes and no responses in Experiment 1 (error bars represent  $\pm 1$  SEM; numbers above bars represent counts of responses)



There was no interaction between Context Rate and Final Word Rate ( $b = 9.34$ ,  $SE = 18.72$ ,  $t = 0.50$ ,  $BF = 0.01$ ). Since the effect of Context Rate was much smaller than the effect of Final Word Rate, we also computed Bayes Factors using a range of narrower priors (in particular, Cauchy priors with a mean of zero and standard deviations of 0.4, 0.6, and 0.8), since wide priors are known to favor large effects over small ones (Rouder et al., 2012). We found weak to strong evidence for an effect of Context Rate ( $BF$  between 2.29 and 4703) and Final Word Rate ( $BF$  between 0.37 and 7778975002) for the other three priors. Importantly, however, evidence against the interaction between Context Rate and Final Word rate was



strong across all priors (BF between 0.02 and 0.67), which is inconsistent the possibility that we simply failed to detect this effect because our priors were too wide.

The lack of interaction between Context Rate and Final Word Rate rules out the possibility that answer times were driven by intelligibility, final word lengthening, or surprise at a rate change. First, if answer times were driven by intelligibility, then we would have expected participants to be slower to answer in the speeded context conditions, where intelligibility was lower, but instead they were slower in the natural context conditions. Second, if answer times were driven by final lengthening, then we would have expected participants to answer earlier in the speeded-natural conditions than all other conditions because the final word was lengthened in comparison to the rest of the utterance. Finally, if answer times were driven by surprise, then we would have expected a larger effect of Final Word Rate after speeded than natural contexts.

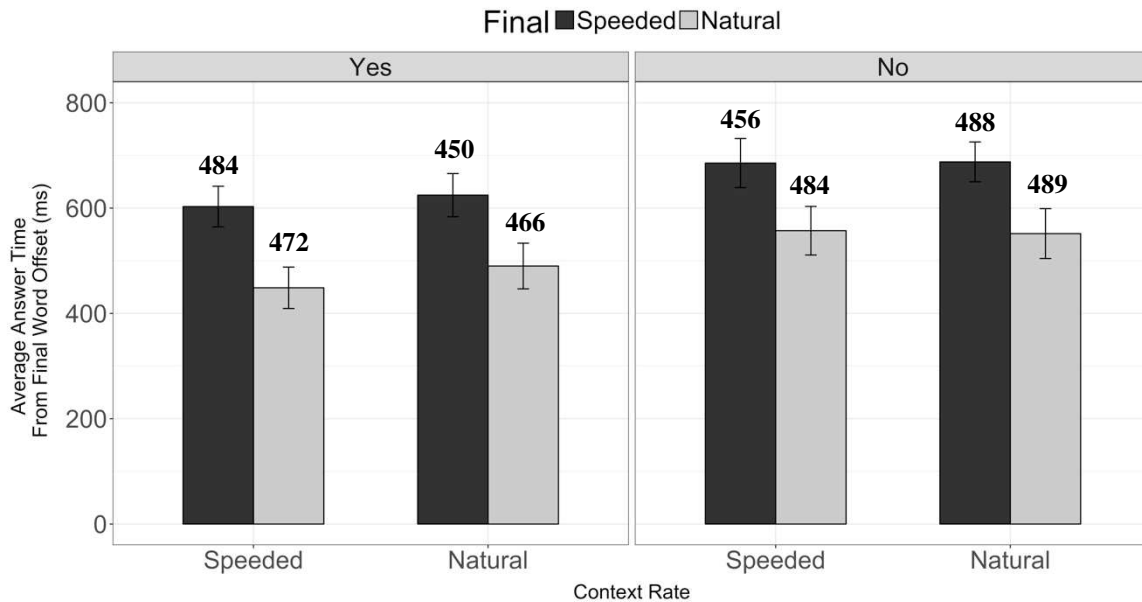
We also found that participants were quicker to answer *yes* than *no* ( $b = -73.90$ ,  $SE = 10.18$ ,  $t = -7.26$ ,  $BF = 116051960$ ; mean answer times for *yes* = 915ms vs. *no* = 998ms), but Answer did not interact with Context Rate ( $b = -27.84$ ,  $SE = 18.31$ ,  $t = -1.52$ ,  $BF = 0.002$ ) or Final Word Rate ( $b = 10.76$ ,  $SE = 18.07$ ,  $t = 0.60$ ,  $BF = 0.02$ ), and there was no three-way interaction ( $b = 34.86$ ,  $SE = 35.81$ ,  $t = 0.97$ ,  $BF = 0.11$ ). Furthermore, participants responded earlier when Agreement was higher ( $b = -27.88$ ,  $SE = 8.70$ ,  $t = -3.20$ ,  $BF = 2.81$ ), but this predictor did not interact with Context Rate ( $b = -1.57$ ,  $SE = 9.92$ ,  $t = -0.16$ ,  $BF = 0.03$ ) or Final Word Rate ( $b = -7.65$ ,  $SE = 9.43$ ,  $t = -0.81$ ,  $BF = 1.64$ ), and there was no three-way interaction ( $b = -34.48$ ,  $SE = 18.65$ ,  $t = -1.85$ ,  $BF = 0.17$ ). Finally, there was no effect of Question Duration ( $b = -18.53$ ,  $SE = 11.79$ ,  $t = -1.52$ ,  $BF = 0.42$ ).

*Analysis from final word offset: Response preparation*

As in the analysis from final word onset, we found that participants answered earlier when context was speeded rather than natural ( $b = -67.44$ ,  $SE = 18.13$ ,  $t = -3.72$ ,  $BF = 2.20$ ; mean answer times for speeded = 572 ms vs. natural = 590 ms; see Figure 2). However, the effect of Final Word Rate was in the opposite direction to that in the analysis from final word onset: Participants answered earlier after a natural than a speeded final word ( $b = 115.79$ ,  $SE = 13.85$ ,  $t = 8.36$ ,  $BF = 367057$ ; mean answer times for natural = 514 ms vs. speeded = 649 ms). As we discussed in the Data Analysis section, this effect most likely occurred because a slow final word gives participants more time to prepare their own verbal response. Therefore, this finding is not informative as to whether the listener adjusted their timing predictions after the rate change, but rather it shows that preparation time has a large effect on the duration of inter-turn intervals. There was no interaction between Context Rate and Final Word Rate ( $b = 8.91$ ,  $SE = 18.89$ ,  $t = 0.47$ ,  $BF = < .001$ ). Note that we found strong evidence for an effect of Context Rate ( $BF$  between 6.44 and 34992) and Final Word Rate ( $BF$  between 871028 and 262340492) across the other three priors, and the evidence against an interaction between Context Rate and Final Word Rate was strong across all priors ( $BF$  between 0 and 0.18), which is inconsistent with the possibility that we simply failed to detect this effect because our priors were too wide.

Figure 2

Average answer times (ms) from final word offset in each of the four conditions for yes and no responses in Experiment 1 (error bars represent  $\pm 1$  SEM; numbers above the bars represent counts of responses)



Again, participants were quicker to answer *yes* than *no* ( $b = -73.50$ ,  $SE = 10.42$ ,  $t = -7.05$ ,  $BF = 15866192$ ; mean answer times for *yes* = 541ms vs. *no* = 620ms), but Answer did not interact with Context Rate ( $b = -28.79$ ,  $SE = 18.31$ ,  $t = 0.47$ ,  $BF = 0.05$ ), Final Word Rate ( $b = 14.83$ ,  $SE = 18.11$ ,  $t = 0.82$ ,  $BF = 1.07$ ), and there was no three-way interaction ( $b = 33.37$ ,  $SE = 35.82$ ,  $t = 0.93$ ,  $BF = 0.25$ ). Answer Agreement was a negative predictor of answer times ( $b = -39.47$ ,  $SE = 7.91$ ,  $t = -4.99$ ,  $BF = 3.43$ ), but did not interact with Context Rate ( $b = -0.64$ ,  $SE = 9.93$ ,  $t = -0.07$ ,  $BF < 0.001$ ), Final Word Rate ( $b = -0.70$ ,  $SE = 9.49$ ,  $t = -0.07$ ,  $BF = 0.001$ ), and there was no three-way interaction ( $b = -35.44$ ,  $SE = 18.67$ ,  $t = -1.90$ ,  $BF = 0.007$ ). But, unlike the analysis from final word onset, Question Duration was a negative predictor: Participants answered earlier when questions were longer in duration ( $b = -35.44$ ,  $SE = 10.99$ ,  $t = -3.23$ ,  $BF = 3.47$ ).

## Discussion

In Experiment 1, we used a *yes/no* question-answering task to investigate how the speech rate of comprehended questions influenced when listeners initiated articulation of their answers during language production. We manipulated the context (e.g., *Do you have a...*) and final word (e.g., *dog?*) rate of questions, so that each component was either produced at a natural or a speeded rate.

When answer times were measured from the onset of the final word, participants answered questions with a speeded context rate earlier compared to those with a natural context rate. This effect is consistent with the shared timing account, but not the separate timing account, and suggests that the context rate of utterances during comprehension influences when listeners launch articulation of their answers during production. Consistent with studies in the speech perception literature (e.g., Dilley & Pitt, 2010), this finding suggests that participants used the context rate of the speaker's question to predict that the speaker would produce the rest of their question at the same rate, allowing them to predict when the speaker would reach the question end and thus the moment they could launch articulation.

We also found that participants answered earlier (again, when measuring from final word onset) when the final word was speeded (speeded-speeded and natural-speeded conditions) rather than natural (natural-natural and speeded-natural conditions). This effect is consistent with the shared timing account and suggests that listeners adjusted their timing predictions immediately after encountering a final syllable that differed in rate from the question, and these predictions immediately affected the timing of subsequent articulation.

Note that, while the effect of context rate was replicated when answer times were analyzed from word offset, the direction of the effect of final word rate was reversed: Participants answered earlier when the final word was natural (and therefore longer in

duration) rather than speeded (and therefore shorter in duration). We interpret this finding as confirming our assumption that participants began response preparation while listening to the speaker's final word. Thus, findings from final word offset reflected the fact that longer words afforded more preparation time, and this preparation advantage was likely so large that it masked the effect of adjusting to final word rate.

However, the effect of context rate did not depend on analysis location in the same way as the final word rate effect. This discrepancy may suggest that the final rate finding is not due to adjusting of timing representations (or else it should behave similarly to the context rate effects), but rather it has an alternative explanation. Perhaps participants respond closer to final word onset when this word is speeded because speeded words are recognized earlier, which in turn allows participants to begin response preparation and launch articulation earlier. In other words, it is possible that our final rate manipulation affects answer times because it affects when participants can begin response preparation rather than because it affects timing predictions. We tested this possibility in Experiment 2 by combining our manipulation of final word rate with a manipulation of content predictability. By making the final word predictable in half of the questions, we allow participants to start preparation before a rate change occurs and before they even hear the final word. Thus, if the final rate effect in Experiment 1 is indeed due to easier recognition, then we should find it is reduced when the final word is predictable.

## **Experiment 2**

In Experiment 1, we found that the timing of when listeners initiated articulation was affected by the rate at which the speaker had produced the majority of their question (i.e., the context rate). Additionally, we found that listeners were affected by the rate of the speaker's final word, and launched articulation earlier when this rate was faster. Although we interpret

these findings as consistent with a shared timing account, and suggest that listeners adjust their timing predictions immediately after encountering a syllable that differed in rate from the rest of the speaker's question, these results are also consistent with the possibility that our speech rate manipulation affected when participants began response preparation. Specifically, listeners may respond closer to the onset of speeded final words because they can recognize them (and the "gist" of the question) earlier, and can thus begin response preparation earlier.

In Experiment 2, we tested this alternative explanation by varying whether response preparation was possible only after recognizing the final word, as in Experiment 1 (unpredictable questions; e.g., *At University, do you study maths?*), or was possible before hearing this word (predictable questions; e.g., *Are dogs your favorite animal?*). Final words in the predictable condition were always consistent with participants' predictions based on context. Thus, since participants could begin preparation before the speaker's final word when listening to predictable questions, the benefit associated with being able to recognize the final word more quickly should be reduced, compared to when the content of the question is unpredictable, and participants can begin preparation only on the speaker's final word. In other words, if the results of Experiment 1 were due to response preparation, we expect an interaction between content predictability and final word rate in Experiment 2, such that the final word rate effect should be reduced for predictable compared to unpredictable questions.

If, however, the final word rate effect in Experiment 1 was due to rapid adjusting of timing representations, then we expect the effect of final word rate to be the same, regardless of the predictability of the final word of the speaker's question. In other words, it should not matter whether preparation can occur early (as in the predictable condition) or not (as in the unpredictable condition): The effect of final word rate should be similar in both cases, because the speech rate manipulation should affect the timing of answer articulation via a shared mechanism, and not indirectly, because it affects the timing of preparation. Thus, we

expect to replicate the final word effect from Experiment 1, and find that listeners respond closer to the onset of a speeded than a natural final word, but crucially we do not expect an interaction between content predictability and final word rate. Of course, we would also expect to replicate previous research and find that listeners respond earlier when content is predictable rather than unpredictable (e.g., Corps et al., 2018).

## **Method**

### *Participants*

Thirty-two additional participants from the same population as in Experiment 1 (9 males;  $M_{age} = 20.10$ ) took part on the same terms. The experiment was approved by the Ethics Committee of the Department of Psychology, University of Edinburgh.

### *Materials*

Using the same norming procedure as in Experiment 1 (22 native English speakers; 6 males;  $M_{age} = 18.5$ ), we elicited completions for 292 question fragments. We assessed the length and content predictability of stimuli and completions using the same procedure as in Experiment 1, but rather than selecting only unpredictable questions, we instead selected 36 predictable content and 36 unpredictable content questions (72 stimuli in total). As intended, stimuli in the predictable content condition had significantly higher content entropy and cloze than those in the unpredictable condition (all  $ps < .001$ ; see Table 3). The predictable and unpredictable conditions were matched for average length entropy ( $p > .17$ ), completion length occurrence ( $p > .46$ ), difficulty, and plausibility (all  $ps > .17$ ; pre-tested with 12 participants, 4 males,  $M_{age} = 18.5$ ).

Table 3: The means (*M*) and standard deviations (*SD*) of content predictability, length predictability, difficulty, plausibility, and intelligibility for stimuli in the predictable and unpredictable conditions of Experiment 2.

Content		Completion Length Entropy <sup>a</sup>	Completion Length Cloze <sup>b</sup>	Completion Content Cloze <sup>c</sup>	Content Entropy <sup>d</sup>	Question Difficulty <sup>e</sup>	Question Plausibility <sup>e</sup>
Predictable	<i>M</i>	0.09	98%	91%	0.46	6.48	6.69
	<i>SD</i>	0.18	5%	9%	0.45	0.62	0.39
Unpredictable	<i>M</i>	0.14	98%	6%	3.09	6.35	6.65
	<i>SD</i>	0.17	3%	3%	0.47	0.49	0.32

<sup>a</sup> Entropy of the length (in number of words) of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion length.

<sup>b</sup> Percentage of participants who provided the word length of the selected completion used in the main experiment (a single word in all conditions) as a continuation in the cloze task.

<sup>c</sup> Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

<sup>d</sup> Entropy of the content of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.



<sup>e</sup> Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, while 7 indicated that the question was very plausible/easy to answer.

Although all final words in Experiment 1 were one syllable long, in Experiment 2 we varied the syllable length of the final word of questions to determine whether the final word effect generalized to multi-syllable items; importantly, we also made sure that the two predictability conditions had the same numbers of one (14), two (14), and three (8) syllable completions. All questions were recorded using the same procedure as Experiment 1. As two of our manipulations were between items (final syllable length and content predictability), it was important to check whether the conditions differed acoustically in any systematic way. Ten (13%) of the questions had creaky voice (four in the predictable condition; six in the unpredictable condition), and so we could not analyze the pitch contours of these stimuli. Among the questions that could be analyzed, all had falling boundary tones and 61 were characterized by a pitch downstep (89% in the predictable condition; 86% in the unpredictable condition). Both judgments were again validated by the same second coder as Experiment 1, who rated 25% of the stimuli and was in perfect agreement with the first coder (Cohen's kappa = 1 for both boundary tone and downstep judgments).

Using the same time-compression method as Experiment 1, we manipulated the rate of the final word of each question (either natural or speeded; see Table 4) and created two versions of each stimulus (natural-natural and natural-speeded). As expected, an ANOVA indicated that final words were significantly longer in the natural than speeded condition ( $F(1, 132) = 327.15, p < .001$ ), and that final words with more syllables had longer spoken durations ( $F(1, 132) = 18.27, p < .001$ ). In addition, final words were longer in the unpredictable than predictable condition ( $F(1, 132) = 7.22, p = 0.01$ ), which may affect our interpretation of effects of content predictability on answer times. If we assume that participants launch articulation when they encounter the word end, then participants will tend to answer questions in the unpredictable condition later than those in the predictable condition simply because they encounter the word end later, rather than because they cannot

begin response preparation early when final words are unpredictable. We return to this issue in the Results section. Importantly, however, there was no three-way interaction between final word rate, syllable length, and content predictability ( $F(1, 132) = 0.13, p = .73$ ), indicating that the rate manipulation was comparable across predictability conditions and the different syllable lengths. All conditions were matched for average intelligibility (all  $ps > .90$ ; mean of 99.9% in all conditions) using the same procedure as Experiment 1.

Table 4: The means (and standard deviations) of the context and final word durations (ms) of questions. The final column provides the difference in the means of the final word durations of the natural and speeded final words.

Content	Final Word Syllable Length	Context Duration	Natural Final Word Duration	Speeded Final Word Duration	Difference in Final Word Duration
Predictable	1	1699 (608)	414 (89)	207 (45)	207
	2	1621 (423)	466 (107)	233 (53)	233
	3	1744 (696)	482 (80)	241 (40)	241
Unpredictable	1	1250 (361)	442 (132)	221 (66)	221
	2	1183 (292)	515 (83)	257 (42)	257
	3	1260 (474)	558 (66)	279 (33)	279

### *Design*

Predictability (predictable vs. unpredictable) was manipulated within participants but between items. Final condition (speeded vs. natural) was manipulated within both

participants and items. We created two stimulus lists (each containing 72 stimuli) using a Latin Square procedure, such that each list contained: (i) 36 predictable and 36 unpredictable stimuli; and (ii) 14 one syllable completions, 14 two syllable completions, and 8 three syllable completions from each of the predictable and unpredictable conditions. Every combination of predictability and final rate condition occurred once across these two lists.

### *Procedure*

The procedure was identical to Experiment 1, except that participants completed 12 practice trials (one from each of the four conditions; one single syllable completion, one two syllable completion, and one three syllable completion) and were given the opportunity to take a break after 36 stimuli.

### **Data Analysis**

Answer times were calculated using the same procedure as Experiment 1. In the unpredictable content conditions, we again expected participants to prepare after final word onset, and so answer times measured from final word offset could be affected by response preparation. In contrast, participants could (in principle) prepare a response before final word onset (i.e., while listening to the context of the speaker's question) in the predictable content conditions. But since we did not manipulate the duration of the context in Experiment 2, the amount of time available for preparation before the final word could not affect answer times measured from final word onset.

We discarded 75 (2.89%) of the 2592 responses because the audio recording was unclear and so the answer could not be categorized as either *yes* or *no*. We then discarded a further three (0.12%) answer times greater than 10000 ms and replaced 33 (1.27%) answer times at the upper limit and 21 (0.81%) at the lower limit. Data analyses, methods, and

predictors were similar to those used in Experiment 1. Thus, we again fitted the full model, in which answer times (from final word onset or offset) were predicted by Content Predictability (reference level: unpredictable vs. predictable), Final Word Rate (reference level: speeded vs. natural), and their interaction. We also included Final Word Syllable Length (and its interactions with Content Predictability and Final Word Rate) as a continuous predictor, since the shared timing account could predict that changes in speech rate during comprehension (e.g., when the speaker suddenly changes their rate of syllable production) do not immediately affect the timing of articulation initiation during language production. Thus, we may expect an interaction between Final Word Rate and Final Word Syllable Length because final words with more syllables offer more time for speech rate changes to affect articulation than final words with fewer syllables.

To account for possible confounding factors, we included interactions between Content Predictability, Final Word Rate, Final Word Syllable Length, Answer, and Answer Agreement. We also included interactions between Content Predictability, Final Word Syllable Length, Answer, Answer Agreement, and Question Duration, but we did not include any interactions between Final Word Rate and Question Duration, since these predictors likely capture some of the same variance: Final words that are speeded will have a shorter duration than those that are natural. In all instances, control predictors were included only in the fixed effects structure, and not as random effects, to aid model convergence. Thus, our random effects structure included random effects for Content Predictability, Final Word Rate, and Final Word Syllable Length by participants and Final Word Rate by items.

All predictors were contrast coded (-0.5, 0.5; where relevant) and centered. Following the same procedure as Experiment 1, we computed Bayes Factors for experimental predictors by fitting *brms* models with Cauchy (0, 1) priors.

## Results and Discussion

### *Analysis from final word onset: Speech rate and articulation initiation*

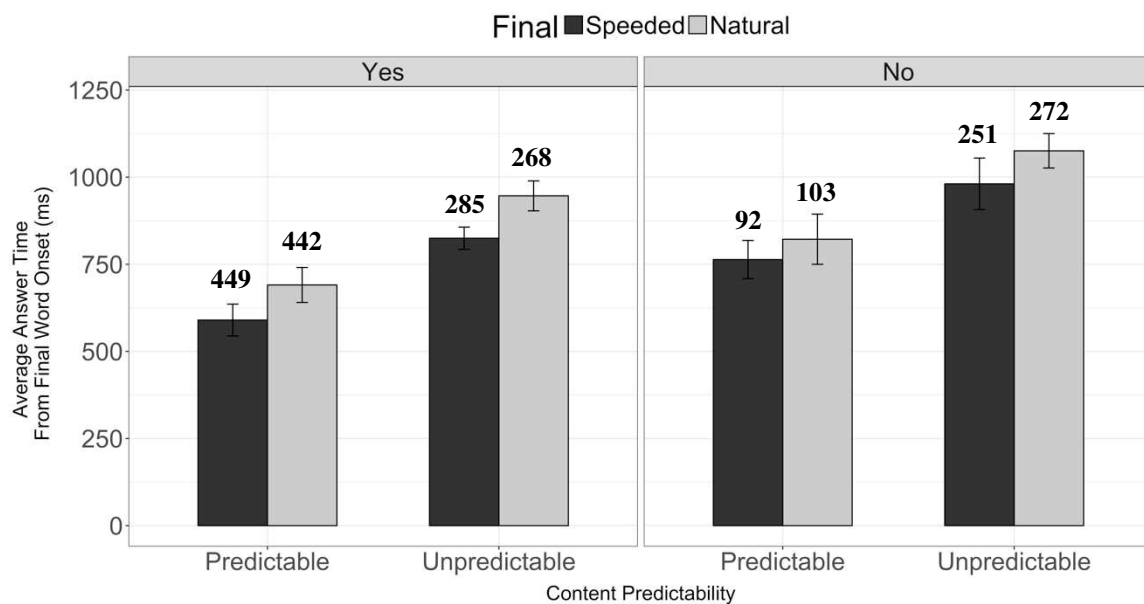
Participants responded earlier when content was predictable than unpredictable ( $b = -204.20$ ,  $SE = 44.13$ ,  $t = -4.63$ ,  $BF = 4431$ ; mean answer times for predictable = 665 ms vs. unpredictable = 947 ms; see Figure 3). Note that the average difference in answer times between the predictable and unpredictable conditions (a difference of 282 ms) is much larger than the average difference in final word duration in these two conditions (a difference of 35 ms), which is inconsistent with the possibility that the effect of content predictability occurred simply because word ends occurred later in the unpredictable than the predictable condition. Instead, this finding suggests that listeners were sensitive to the content predictability of the speaker's question and used this information to prepare a response as early as possible (e.g., Bögers et al., 2015; Corps et al., 2018).

We also found a significant effect of Final Word Rate: Participants responded earlier after a speeded than a natural final word ( $b = -111.78$ ,  $SE = 24.28$ ,  $t = -4.60$ ,  $BF = 200164$ ; mean answer times for speeded = 748 ms vs. natural = 865 ms), thus replicating Experiment 1. There was no interaction between Content Predictability and Final Word Rate ( $b = 27.09$ ,  $SE = 43.90$ ,  $t = 0.62$ ,  $BF = 0.08$ ), suggesting that our effect of final word rate in Experiment 1 did not occur simply because participants recognized the speaker's final word and began preparation earlier in the speeded than the natural condition. Note that we report Bayes Factors for models fitted with Cauchy(0, 1) priors, but we found strong evidence for an effect of Content Predictability ( $BF$  between 3689 and 5720) and Final Word Rate ( $BF$  between 9705 and 406633) for all four priors. Importantly, evidence against the interaction between Content Predictability and Final Word Rate was strong across all priors ( $BF$  between 0.10 and 0.14).

Furthermore, the number of syllables did not influence answer times ( $b = -4.17$ ,  $SE = 18.02$ ,  $t = -0.23$ ,  $BF = 0.02$ ) and did not interact with Final Word Rate ( $b = -15.17$ ,  $SE = 25.69$ ,  $t = -0.59$ ,  $BF = 0.05$ ) or Content Predictability ( $b = -16.77$ ,  $SE = 36.27$ ,  $t = -0.46$ ,  $BF = 0.09$ ), and there was no three-way interaction among these predictors ( $b = -15.52$ ,  $SE = 51.15$ ,  $t = -0.31$ ,  $BF = 0.09$ ). Thus, our final word effect from Experiment 1 generalized to multi-syllabic words.

Figure 3

Average answer times (ms) from final word onset in each of the four conditions for yes and no responses in Experiment 2 (error bars represent  $\pm 1$  SEM; numbers above the bars represent counts of responses)



As in Experiment 1, participants were quicker to answer *yes* than *no* ( $b = -94.52$ ,  $SE = 33.88$ ,  $t = -2.79$ ,  $BF = 4.26$ ). Unlike in Experiment 1, Answer interacted with Final Word Rate ( $b = -101.32$ ,  $SE = 49.53$ ,  $t = -2.05$ ,  $BF = 1.10$ ). We followed-up this interaction by fitting separate models to *yes* and *no* responses, using the same analysis structure as the full models, but without Answer included as a predictor. Importantly, participants answered earlier after a speeded final word for both types of response (see Figure 3), even though Final

Word Rate was a stronger predictor of answer times for *yes* ( $b = -135.80$ ,  $SE = 21.24$ ,  $t = -6.39$ ,  $BF = 49217$ ) compared to *no* ( $b = -81.77$ ,  $SE = 37.03$ ,  $t = -2.21$ ,  $BF = 0.14$ ) answers. Moreover, there was no evidence for either the three-way interaction among Content Predictability, Final Word Rate, and Answer ( $b = 85.95$ ,  $SE = 98.99$ ,  $t = 0.87$ ,  $BF = 0.24$ ) or the two-way interaction between Answer and Content Predictability ( $b = 66.70$ ,  $SE = 67.81$ ,  $t = 0.98$ ,  $BF = 0.18$ ). In sum, across both *yes* and *no* answers, there was no indication that the magnitude of the Final Word Rate effect varied with Predictability (see Figure 3).

As in Experiment 1, Answer Agreement was a negative predictor of answer times ( $b = -45.08$ ,  $SE = 17.85$ ,  $t = -2.53$ ,  $BF = 0.47$ ), but did not interact with Content Predictability ( $b = -28.56$ ,  $SE = 35.72$ ,  $t = -0.80$ ,  $BF = 0.10$ ) or Final Word Rate ( $b = 7.58$ ,  $SE = 21.83$ ,  $t = 0.35$ ,  $BF = 0.03$ ), and there was no three-way interaction ( $b = 46.92$ ,  $SE = 43.68$ ,  $t = 1.07$ ,  $BF = 0.20$ ). Question Duration again did not predict answer times ( $b = -11.22$ ,  $SE = 20.99$ ,  $t = -0.53$ ,  $BF = 0.06$ ) and did not interact with Content Predictability ( $b = 59.30$ ,  $SE = 42.02$ ,  $t = 1.41$ ,  $BF = 0.10$ ).

#### *Analysis from final word offset: Response preparation*

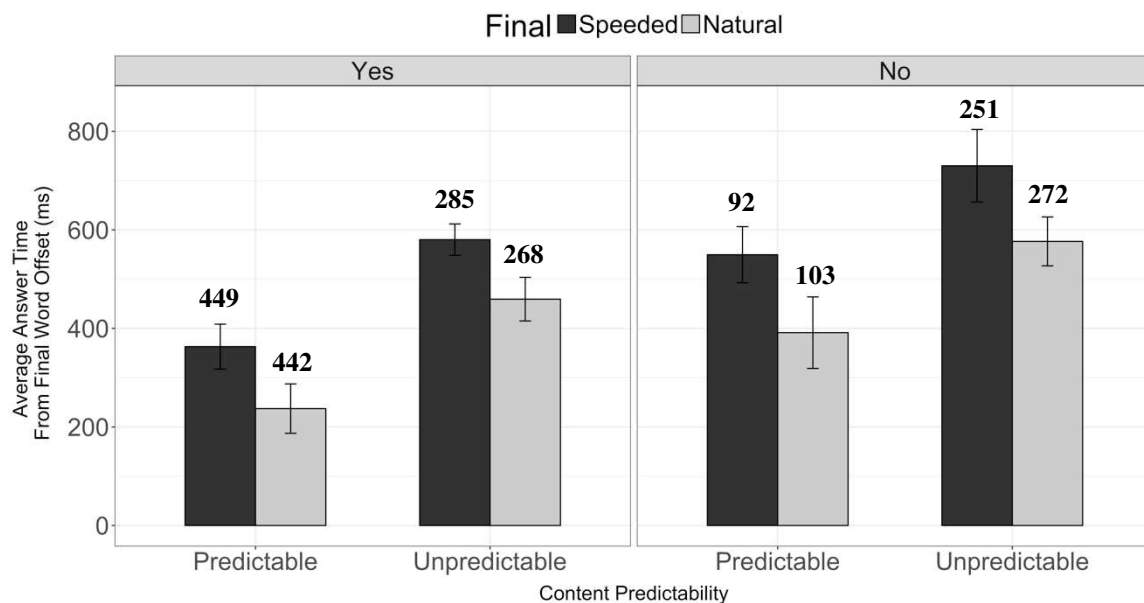
In our analysis from final word offset, we replicated the finding that participants answered earlier when questions were predictable rather than unpredictable in content ( $b = -159.41$ ,  $SE = 41.16$ ,  $t = -3.87$ ,  $BF = 313$ ; mean answer times for predictable = 328 ms vs. unpredictable = 577 ms; Figure 4). As in our analysis from final word offset in Experiment 1, participants answered earlier when the final word was natural rather than speeded ( $b = 116.80$ ,  $SE = 24.06$ ,  $t = 4.85$ ,  $BF = 912$ ; mean answer times for natural = 393 ms vs. speeded = 511 ms). Crucially, however, there was no interaction between these two factors ( $b = -16.96$ ,  $SE = 43.06$ ,  $t = -0.35$ ,  $BF = 0.12$ ). In addition, Final Word Syllable Length was not a significant predictor ( $b = -30.87$ ,  $SE = 16.50$ ,  $t = -1.87$ ,  $BF = 0.12$ ), and there was no



interaction between Final Word Syllable Length and Content Predictability ( $b = -18.61$ ,  $SE = 33.25$ ,  $b = -0.56$ ,  $BF = 0.05$ ) or between Final Word Syllable Length and Final Word Rate ( $b = -1.16$ ,  $SE = 25.53$ ,  $t = -0.05$ ,  $BF = 0.07$ ), and no three-way interaction among Final Word Syllable Length, Content Predictability, and Final Word Rate ( $b = -22.26$ ,  $SE = 50.83$ ,  $t = -0.44$ ,  $BF = 0.09$ ).

Figure 4

Average answer times (ms) from final word offset in each of the four conditions for no and yes responses (error bars represent  $\pm 1$  SEM; numbers above the bars represent counts of responses)



Participants answered *yes* faster than *no* ( $b = -105.53$ ,  $SE = 32.68$ ,  $t = -3.23$ ,  $BF = 9.86$ ; mean answer times for *yes* = 387ms vs. *no* = 585ms). As in the analysis from final word onset, there was no three-way interaction among Content Predictability, Final Word Rate, and Answer ( $b = 102.95$ ,  $SE = 98.28$ ,  $t = 1.05$ ,  $BF = 0.02$ ) and no two-way interaction between

Content Predictability and Answer ( $b = 53.87$ ,  $SE = 65.43$ ,  $t = 0.82$ ,  $BF = 0.03$ ). There was, however, a marginally significant interaction between Final Word Rate and Answer ( $b = -93.70$ ,  $SE = 49.17$ ,  $t = -1.91$ ,  $BF = 1.20$ ). Following-up this interaction, using the same procedure as the analysis of final word onset, we found that Final Word Rate was a stronger predictor for *yes* ( $b = 92.38$ ,  $SE = 21.19$ ,  $t = 4.36$ ,  $BF = 16.77$ ) than *no* ( $b = 143.50$ ,  $SE = 37.13$ ,  $t = 3.87$ ,  $BF = 72.07$ ) responses, but crucially Final Word Rate was a significant predictor for both types of responses.

Answer Agreement was a negative predictor of response times ( $b = -53.08$ ,  $SE = 16.10$ ,  $t = -3.30$ ,  $BF = 3.26$ ), but did not interact with Content Predictability ( $b = -43.58$ ,  $SE = 32.21$ ,  $t = -1.35$ ,  $BF = 0.21$ ) or Final Word Rate ( $b = 21.17$ ,  $SE = 21.36$ ,  $t = 0.99$ ,  $BF = 0.08$ ) and there was no three-way interaction ( $b = 49.63$ ,  $SE = 42.75$ ,  $t = 1.16$ ,  $BF = 0.17$ ). Unlike the analysis from final word offset in Experiment 1, Question Duration did not predict answer times ( $b = -13.91$ ,  $SE = 19.06$ ,  $t = -0.73$ ,  $BF = 0.06$ ), and did not interact with Content Predictability ( $b = 25.52$ ,  $SE = 38.14$ ,  $t = 0.67$ ,  $BF = 0.11$ ).

## General Discussion

In two experiments, we used a verbal question-answering task to investigate whether the speech rate of utterances during language comprehension influenced when participants initiated answer articulation. We contrasted two accounts: (1) a shared timing account, which assumes that the speech rate of utterances during comprehension can influence the timing of articulation initiation during language production, possibly over multiple timescales (specifically, a single utterance and a single syllable); and (2) a separate timing account, which assumes that the speech rate of utterances during comprehension should not influence the timing of articulation initiation during language production. To do so, we combined a

natural or speeded context with a natural or speeded final word (in Experiment 1; in Experiment 2, the context was always natural and only the final word rate was manipulated).

Crucially, we found that the timing of answer initiation was influenced by the speech rate of questions, providing clear support for a shared timing account. In particular, in Experiment 1 we found that participants' answer times, from both final word onset and offset, were influenced by the context rate of questions: They answered earlier when the context was speeded (twice as fast as its original rate) rather than natural. Note that this effect was consistently present across analyses from both final word onset and final word offset, which suggests it is likely to affect the timing of inter-turn intervals regardless of the duration of the final word of turns. Although this effect was small (19 ms difference on average between speeded and natural conditions from final word onset; 18 ms difference from final word offset), it is consistent with evidence of speech rate priming (e.g., Jungers & Hupp, 2009) and extends these findings to suggest that the rate of the speaker's turn influences not only the rate of the listener's (as the next speaker) own response, but also the timing of its initiation in relation to the end of the speaker's utterance.

In addition to this context effect, we found that listeners responded earlier (when measuring from final word onset) when the speaker's final syllable was speeded rather than natural, regardless of context rate. Although this finding is also consistent with a shared timing account, and suggest that listeners adjust their timing predictions immediately after encountering a syllable differing in rate from the rest of a speaker's utterance, it may also be explained by ease of recognition: When a final word is speeded, participants can recognize the word and thus begin response preparation earlier than when the final word is natural.

We tested this possibility in Experiment 2, in which we varied whether response preparation was possible only after recognizing the final word (unpredictable condition; e.g., *At University, do you study maths?*), or was possible before hearing this word (predictable

questions; e.g., *Are dogs your favorite animal?*). Participants answered earlier when the content of the speaker's final word was predictable rather than unpredictable, which is consistent with previous question-answering tasks (e.g., Bögels et al., 2015; Corps et al., 2018) and indicates that listeners used content predictions to prepare a verbal response. But importantly we replicated our final word rate effect from Experiment 1 and found no evidence that this final word rate effect was influenced by content predictability, which is inconsistent with the possibility that participants in Experiment 1 responded earlier when the final word was speeded rather than natural simply because the disambiguating information necessary for recognizing the speaker's final word occurred earlier (and subsequent response preparation could thus also occur earlier).

Together, these findings are consistent with studies demonstrating that listeners use the speech rate of utterances over multiple timescales (e.g., a single utterance and multiple utterances) to predict the rate of forthcoming speech (e.g., Baese-Berk et al., 2014; Dilley & Pitt, 2010). In other words, separate effects of context and final word rate in Experiment 1 suggest that listeners form and sustain timing predictions over long timescales (i.e., multiple syllables), but can also adjust their predictions rapidly over shorter timescales (i.e., a single syllable). Crucially, our experiments demonstrate that timing predictions over multiple timescales during comprehension can influence the timing of later production, which is consistent with a shared timing account.

But how can listeners integrate speech rate over multiple time-scales? Some theories suggest that neural oscillators in the auditory cortex can track the speech stream at different frequencies, which correspond to different linguistic units. For example, Giraud and Poeppel (2012) suggest that oscillations in the theta range (4-8 Hz) correspond to syllabic processing, while those in the delta range (1-3 Hz) correspond to phrasal processing. Thus, in our experiments, these oscillations may have been sensitive to both the context and the final word

rate of our questions. But an additional mechanism is needed to explain how entrainment in auditory cortex over multiple time-scales can in turn affect articulation. Interestingly, research suggests that entrainment between speech and auditory cortex in both the theta and the delta range is modulated by coupling between auditory and pre-motor areas of the brain (e.g., Park et al., 2015), which would explain why articulation was affected by speech rate over multiple time-scales in our experiments.

In sum, our findings demonstrate that the timing of articulation during language production is influenced by speech rate, even though the precise mechanism responsible for this influence is yet to be determined. Such findings are consistent with theoretical accounts of conversational turn-taking, which argue that speech rate plays a role in coordinating turns (e.g., Wilson & Wilson, 2005; Garrod & Pickering, 2015). Although further research is needed to investigate this issue, listeners could coordinate their utterances by using speech rate in parallel with other mechanisms that have been shown to be involved in timing response articulation, such as turn-end prediction (e.g., De Ruiter et al., 2006) and turn-final cues (e.g., Gravano & Hirschberg, 2011). For example, if the listener predicts that the speaker will soon produce their final word (a turn-end prediction), then they could use the speech rate of the speaker's utterance to predict how long it is likely to take the speaker to produce their final word, and thus the moment that articulation can be launched.

In Experiment 2, our findings demonstrated that our speech rate manipulation affects response timing independently of response preparation. Although further research is needed to confirm this finding, since our conclusions are based on a null interaction, we note that this result is consistent with Bögels and Levinson's (2017; see also Levinson & Torreira, 2015) early-planning hypothesis, which claims that listeners often prepare their verbal response independently from timing articulation (i.e., without necessarily knowing when they will have the opportunity to launch articulation). On occasions when listeners do prepare early,

they must hold this response in an articulatory buffer until they can launch articulation (Piai, Roelofs, Rommers, Dahlsätt, & Maris, 2015). Conversely, there may also be instances when listeners know they can begin articulation, but have not yet prepared their response. In these instances, the listener most likely has to plan later aspects of their response incrementally at the same time as they articulate earlier aspects of their response (e.g., Ferreira & Swets, 2002).

Although we focused on the timing of participants' responses (i.e., how quickly they answered), we also note that faster responses are not necessarily better. Interlocutors need not only ensure they produce their response quickly, but they must also do so without extensively overlapping with the previous speaker, in part because conversational overlap may reduce intelligibility. In other words, listeners must ensure they produce their response both *quickly* and *precisely*. Indeed, in one previous study, we have considered both the timing and the precision (i.e., how closely participants respond to the end of the speaker's turn) of responses (Corps et al., 2018). We did not discuss response precision in this work because the accounts we test make specific predictions about timing and not precision. In other words, if the speech rate of utterances during language comprehension can influence later production, then we expect a listener comprehending a turn at a speeded rate to respond faster, but we do not make any predictions regarding answer precision.<sup>1</sup>

In our experiments, we showed that listening to a speeded speech rate leads listeners to *begin* responding faster. Previous research indicates that a speaker's speech rate is primed by the rate of a previous speaker's utterance (e.g., Jungers & Hupp, 2009), so it is also possible that our manipulations affected articulation rate as well as the timing of articulation initiation. We did not test this possibility in the current study because participants responded with short *yes* or *no* answers, which are not ideal for investigating articulation rate. However,

---

<sup>1</sup> The interested reader can find model outputs for precision analysis at: <https://osf.io/4bt3/>.

future research could combine our design with longer answers to investigate how rapidly speech rate changes during comprehension affect articulation rate during production, given that previous studies (e.g., Jungers & Hupp, 2009; Jungers et al., 2002) have focused on long-term changes (across many utterances) and have not investigated the effect of more short-term changes (e.g., on the final syllable).

Finally, although our intent in using a question-answering task was to investigate the mechanisms that underlie conversational turn-taking, it is worth noting that inter-turn intervals in our experiments (i.e., answer times from final word offset; mean of 583, mode of 713 ms for Experiment 1; mean of 453 ms, mode of 245 ms for Experiment 2) are longer than those displayed in natural conversation (mode of around 200 ms; Stivers et al., 2009). It is likely interlocutors time their utterances using whatever cues are available during dialogue, and the absence of much of this information may mean that our task elicits slower response times than those in natural dialogue. For example, even though our questions were easy, participants still had to provide the correct answer and research suggests that a quiz-like setting often elicits slower response times than natural conversation (e.g., Bögels et al., 2015). Furthermore, interlocutors in natural conversation can often use the surrounding conversational context to constrain their interpretation of the current utterance (e.g., Corps et al., 2018), which could aid prediction. We propose that our task can be successfully used to investigate the mechanisms that may underlie conversational turn-taking (e.g., Corps et al., 2018) because it represents a simplified, but accurate model of this natural phenomenon. Nevertheless, future research should assess whether our task can be modified to more closely approximate the timing of conversational turn-taking (e.g., Meyer, Alday, Decuyper, & Knudsen, 2018).

In conclusion, we have shown that participants in a question-answering task use speech rate over multiple timescales (a single utterance and a single syllable) to time

response articulation, suggesting that the speech rate of utterances during language comprehension can influence the timing of articulation initiation during language production. In addition, we found that our speech rate manipulation did not affect the process of response preparation, and thereby argued that the processes involved in response preparation and articulation often occur independently during conversational turn-taking.



## References

- Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, 53, 741-764.
- Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16, 390-398.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, 25, 1546-1553.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed-effects models using "Eigen" and S4* (R package version 1.1-8). Retrieved from <http://CRAN.R-project.org/package=lme4>.
- Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255-309.
- Beňuš, S. (2009). Are we 'in sync': Turn-taking in collaborative dialogues. In *INTERSPEECH 2009 – 10<sup>th</sup> Annual Conference of the International Speech Communication Association*.
- Boersma, P. & Weenink, D. (2002). Praat: Doing phonetics by computer (version 6.0.17) [Computer software]. Retrieved June 20, 2017, from <http://www.praat.org>.

- Bögels, S. & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, 50, 71-89.
- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5, <http://dx.doi.org/10.1039/srep12881>.
- Bögels, S. & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends. *Journal of Phonetics*, 52, 46-57.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75, B13-B25.
- Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482-1493.
- Bürkner, P-C. (2017). *Brms: Bayesian Regression Models using Stan* (R package version 1.6.1). Retrieved from <https://CRAN.R-project.org/package=brms>
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition*, 175, 77-95.
- Corps, R. E., Pickering, M. J., Gambi, C. (2018). Predicting turn-ends in discourse context. *Language, Cognition and Neuroscience*, DOI: 10.1080/23273798.2018.1552008.
- Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology*, 3, <https://dx.doi.org/10.3389/fpsyg.2012.00364>.
- Cutler, A. & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation and Discourse*. (pp. 139-155). London: Croom Helm.

- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222-241.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82, 515-535.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, <https://dx.doi.org/10.3389/fpsyg.2014.00781>.
- Dilley, L. C., Morrill, T. H., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalization. *Frontiers in Psychology*, 4, <http://dx.doi.org/10.3389/fpsyg.2013.01002>.
- Dilley, L. C. & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21, 1664-1670.
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81, 181-187.
- Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, 761-768.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, 23, 283-292.
- Ferreira, F. & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57-84.

- Finlayson, I., Lickley, R. J., & Corley, M. (2012, March). *Convergence of speech rate: Interactive alignment beyond representation*. Paper presented at the 25<sup>th</sup> Annual CUNY Conference on Sentence Processing, New York.
- Gambi, C. & Pickering, M. J. (2017). Models linking production and comprehension. In E. M. Fernández & H. Smith Cairns (Eds.), *The Handbook of Psycholinguistics*. (pp. 157-181). Oxford: John Wiley & Sons.
- Garrod, S. & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6, <http://dx.doi.org/10.3389/fpsyg.2015.00751>.
- Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, 3, <http://dx.doi.org/10.3389/fpsyg.2012.00238>.
- Giraud, A. L. & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511-517.
- Gravano, A. & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, 25, 601-634.
- Jungers, M. K. & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes*, 24, 611-624.
- Jungers, M. K., Palmer, C., & Speer, S. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, 1, 21-35.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation*, 54, 59-101.
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2017). Neural entrainment determines the words we hear. *BioRxiv*, <http://dx.doi.org/10.1101/175000>.

- Large, E. W. & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119-159.
- Levinson, S. C. & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, <http://dx.doi.org/10.3389/fpsyg.2015.0073>.
- Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 224-234.
- Magyari, L. & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3, <http://dx.doi.org/10.3389/fpsyg.2012.00376>.
- Magyari, L., De Ruiter, J. P., Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, 8, <https://dx.doi.org/fpsyg.2017.00211>
- Matuschek, H., Kliegel, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type 1 error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315.
- Menenti, L., Gierham, S. M., Segaert, K., & Hagoort, P. (2011). Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, 22, 1173-1182.
- Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, 9, <https://dx.doi.org/10.3389/fpsyg.2018.00525>.

- Miller, J. L. & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 369-378.
- Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41, 215-225.
- Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition*, 131, 69-74.
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5-6), 453-467.
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, 25, 1649-1653.
- Pecenka, N. & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211, 505-515.
- Peelle, J. E. & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, 3, <https://dx.doi.org/10.3389/fpsyg.2012.00320>.
- Piai, V., Roelofs, A., Rommers, J., Dahlsätt, K., & Maris, E. (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience*, 9, <https://doi.org/10.3389/fnhum.2015.00549>.
- Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context rate. *Attention, Perception, & Psychophysics*, 78, 334-345.

- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356-374.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86-102.
- Schultz, B. G., O'Brien, I., Phillips, N., & McFarland, D. H. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, 37, 1201-1220.
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2011). Shared syntax in language production and language comprehension – an fMRI study. *Cerebral Cortex*, 22, 1662-1670.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 76, 379-423.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106, 10587-10592.
- Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, 11, 139-169.
- Strömbergsson, S., Hjalmarsson, A., Edlund, J., & House, D. (2013). Timing responses to questions in dialogue. In F. Bimbot. *INTERSPEECH 2013*. Paper presented at 14<sup>th</sup> Annual Conference of the International Speech Communication Association (pp. 2584-2588). Lyon: International Speech and Communication Association.
- Tauroza, S. & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, 11, 90-105.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.

Wilson, M. & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking.

*Psychonomic Bulletin & Review*, 12, 957-968.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M. . .

. & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, 77, 980-991

..



Appendix A: Linear mixed effects model output for the answer times analysis from final word onset and final word offset for both experiments

Table A1: Fixed (top) and random (bottom) effects structure for the lmer analysis of answer times from final word onset (left) and final word offset (right) in Experiment 1. All predictors are defined in the Data Analysis section.

	Answer times from final word onset				Answer times from final word offset			
Fixed effect	Coefficient	SE	t	BF	Coefficient	SE	t	BF
Intercept	956.47	40.67	23.52		581.82	40.57	14.34	
Context Rate (Slow – Fast)	-44.35	19.07	-2.33	295	-67.44	18.13	-3.12	2.20
Final Word Rate (Slow – Fast)	-122.95	13.74	-8.95	95721635	115.79	13.85	8.36	367057
Context Rate * Final Word Rate	9.34	18.72	0.50	0.01	8.91	18.89	0.47	15866192
Question Duration	-18.53	11.79	-1.57	0.42	-35.44	10.99	-3.23	3.47
Answer (No – Yes)	-73.66	10.16	-6.94	116051960	-73.05	10.42	-7.05	15866192
Answer * Context Rate	-27.84	18.31	-1.52	0.002	-28.79	18.31	-1.57	0.05
Answer * Final Word Rate	10.78	18.07	0.60	0.02	14.83	18.11	0.82	1.07
Answer * Context Rate * Final Word Rate	34.86	35.81	0.97	0.11	33.37	35.82	0.93	0.25

Answer Agreement	-27.77	8.70	-3.19	2.81	-39.47	7.91	-4.99	3.43
Answer Agreement * Context	-1.57	9.92	-0.16	0.03	-0.64	9.93	-0.07	< .001
Rate								
Answer Agreement * Final Word	-7.65	9.43	-0.81	1.64	-0.70	9.49	-0.07	0.001
Rate								
Answer Agreement * Answer	2.22	10.81	0.21	0.43	2.79	10.59	0.26	< .001
Answer Agreement * Context	-34.48	18.65	-1.85	0.17	-35.44	18.67	-1.90	0.007
Rate * Final Word Rate								
Answer Agreement * Answer *	1.38	18.47	0.08	0.30	0.99	18.47	0.05	0.007
Context Rate								
Answer Agreement * Answer *	-34.48	18.65	-1.85	0.67	-1.80	18.18	-0.10	0.01
Final Word Rate								
Answer Agreement * Answer *	27.63	36.03	0.78	0.05	25.30	36.04	0.70	0.004
Context Rate * Final Word Rate								
Random effect		Variance	SD			Variance	SD	
Item (Intercept)		6995	83.64			5367	73.26	

Item (Context Rate)	2633	51.31	2699	51.95
Item (Final Word Rate)	1592	39.90	1754	41.88
Item (Context Rate * Final Word Rate)	5294	72.76	5498	74.15
Participant (Intercept)	50520	224.80	50673	225.11
Participant (Context Rate)	0	0.00	0	0.00
Participant (Final Word Rate)	2071	45.51	2272	47.67
Participant (Context Rate * Final Word Rate)	289	16.99	446	21.11

---

Table A2: Fixed (top) and random (bottom) effects structure for the lmer analysis of answer times from final word onset (left) and final word offset (right) in Experiment 2. All predictors are defined in the Data Analysis section.

Fixed effect	Answer times from final word onset				Answer times from final word offset			
	Coefficient	SE	t	BF	Coefficient	SE	t	BF
Intercept	808.50	42.14	19.19		465.98	41.51	11.23	
Content Predictability (Unpredictable - Predictable)	-204.20	44.13	-4.63	4431	-159.41	41.16	-3.87	313
Final Word Rate (Slow – Fast)	-111.78	24.28	-4.60	200164	116.80	24.06	4.85	912
Content Predictability * Final Word Rate	27.09	43.90	0.62	0.08	-16.96	43.06	-0.39	0.12
Syllable length	-4.17	18.02	-0.23	0.02	-30.87	16.50	-1.87	0.12
Syllable length * Content Predictability	-16.77	36.27	-0.46	0.09	-18.61	33.25	-0.56	0.07
Syllable length * Final Word Rate	-15.17	25.69	-0.59	0.05	-1.16	25.53	-0.05	0.05
Syllable length * Content Predictability * Final Word Rate	-15.62	51.15	-0.31	0.09	-22.28	50.83	-0.44	0.09
Question Duration	-11.22	20.99	-0.53	0.06	-13.91	19.06	-0.73	0.06

Question Duration * Content Predictability	59.30	42.02	1.41	0.10	25.52	38.14	0.67	0.11
Answer (No – Yes)	-94.52	33.88	-2.79	4.26	-105.53	32.68	-3.23	9.86
Answer * Content Predictability	66.70	67.81	0.98	0.18	53.97	65.43	0.82	0.03
Answer * Final Word Rate	-101.32	49.53	-2.05	1.10	-93.70	49.17	-1.91	1.07
Answer * Content Predictability * Final	85.95	98.99	0.87	0.24	102.95	98.27	1.05	0.02
Word Rate								
Answer * Syllable length	46.33	36.70	1.26	0.09	47.47	35.78	1.32	0.12
Answer * Syllable length * Content	68.15	73.19	0.93	0.24	81.90	71.34	1.15	0.26
Predictability								
Answer * Syllable length * Final Word Rate	-46.08	66.74	-0.69	0.13	-32.08	66.26	-0.48	0.10
Answer * Syllable length * Content	113.78	133.09	1.01	0.28	145.86	132.12	1.10	0.24
Predictability * Final Word Rate								
Answer * Question Duration	-97.61	35.30	-2.77	14.36	-86.92	34.01	-2.56	7.07
Answer * Question Duration * Content	79.18	70.79	1.12	0.32	88.15	68.18	1.29	0.57
Predictability								
Answer Agreement	-45.08	17.85	-2.53	0.46	-53.08	16.10	-3.30	3.26

Answer Agreement * Content Predictability	-28.56	35.72	-0.80	0.10	-43.58	32.21	-1.35	0.21
Answer Agreement * Final Word Rate	7.58	21.83	0.35	0.03	21.17	21.36	0.99	0.08
Answer Agreement * Content Predictability *	46.92	43.68	1.07	0.20	49.63	42.75	1.16	0.17
Final Word Rate								
Answer Agreement * Syllable length	-36.32	16.80	-2.16	0.17	-24.10	15.26	-1.58	0.08
Answer Agreement * Syllable length *	-7.96	33.55	-0.24	0.05	-22.30	30.45	-0.73	0.07
Content Predictability								
Answer Agreement * Syllable length * Final	10.39	22.90	0.45	0.05	-1.10	22.77	-0.05	0.05
Word Rate								
Answer Agreement * Syllable length *	-110.47	45.59	-2.42	1.04	-04.41	45.33	-2.30	0.72
Content Predictability * Final Word Rate								
Answer Agreement * Answer	-7.18	31.01	-0.23	0.04	-16.04	30.21	-0.51	0.08
Answer Agreement * Answer * Content	15.26	62.01	0.25	0.11	-3.22	60.43	-0.05	0.12
Predictability								
Answer Agreement * Answer * Final Word	-117.04	47.47	-2.47	1.71	-1.10	22.77	-0.05	1.11
Rate								

Answer Agreement * Answer * Content Predictability * Final Word Rate	87.43	94.64	0.92	0.20	103.97	94.01	1.11	0.20
Answer Agreement * Answer * Syllable length	51.92	30.49	1.70	0.09	52.18	29.79	1.77	0.12
Answer Agreement * Answer * Syllable length * Content Predictability	-34.69	60.91	-0.57	0.09	-20.26	59.49	-0.34	0.10
Answer Agreement * Answer * Syllable length * Final Word Rate	29.66	56.45	0.53	0.09	38.37	56.06	0.68	0.13
Answer Agreement * Answer * Syllable length * Content Predictability * Final Word Rate	172.57	112.75	1.53	0.33	176.49	111.98	1.58	0.12
Answer Agreement * Question Duration	-18.56	21.98	-0.84	0.05	-2.62	19.96	-0.13	0.04
Answer Agreement * Question Duration * Content Predictability	4.31	44.04	0.10	0.08	1.29	39.98	0.03	0.51
Answer Agreement * Question Duration * Answer	-88.98	36.31	-2.45	2.45	-82.23	35.26	-2.33	1.40

Answer Agreement \* Question Duration \*      -73.82      72.71      -1.02      0.12      -87.74      70.59      -1.24      0.23

Answer \* Content Predictability

Random effect	Variance	<i>SD</i>	Variance	<i>SD</i>
Item (Intercept)	8298	91.09	5680	75.36
Item (Final Word Rate)	0	0.00	0	0.00
Participant (Intercept)	46414	215.44	46670	216.00
Participant (Content Predictability)	20645	143.68	20280	142.40
Participant (Final Word Rate)	3428	58.55	3669	60.57
Participant (Syllable length)	0	0.00	0	0.00
Participant (Content Predictability * Final Word Rate)	0	0.00	0	0.00
Participant (Content Predictability * Syllable length)	638	25.25	640	25.29
Participant (Final Word Rate * Syllable length)	3428	58.55	0	0.00



Participant (Content Predictability * Final Word Rate * Syllable length)	0	0.00	0	0.00
--	---	------	---	------

Table A3: Fixed (top) and random (bottom) effects structure for the lmer analysis of answer times from final word onset. Note that models were fitted separately for *yes* (left) and *no* (right) responses to follow-up the two-way interaction between Final Word Rate and Answer when analyzing answer times from final word onset in Experiment 2 (see Table A3)

	Yes responses			No responses		
Fixed effect	Coefficient	SE	t	Coefficient	SE	t
Intercept	737.82	43.86	16.82	946.26	48.15	19.65
Content Predictability (Unpredictable - Predictable)	-173.82	40.84	-4.26	-217.20	56.30	-3.86
Final Word Rate (Slow – Fast)	-141.11	22.14	-6.37	-80.48	37.77	-2.13
Content Predictability * Final Word Rate	66.26	45.65	1.45	-25.77	76.50	-0.34
Syllable length	3.18	17.25	0.18	21.01	23.06	0.91
Syllable length * Content Predictability	12.40	37.28	0.33	-61.55	54.48	-1.13
Syllable length * Final Word Rate	-20.64	19.72	-1.05	26.31	36.70	0.72
Syllable length * Content Predictability * Final Word Rate	12.81	42.23	0.30	58.52	94.99	0.62

Question Duration	-47.97	21.88	-2.19	28.04	26.31	1.07
Question Duration * Content Predictability	76.49	51.36	1.49	-2.93	46.91	-0.06
Answer Agreement	-59.74	17.62	-3.39	-45.23	25.65	-1.76
Answer Agreement * Content Predictability	-23.24	36.25	-0.64	-53.40	50.07	-1.07
Answer Agreement * Final Word Rate	-26.78	21.99	-1.22	74.69	37.98	1.97
Answer Agreement * Content Predictability * Final Word Rate	71.70	43.70	1.64	-7.11	79.64	-0.09
Answer Agreement * Syllable length	-22.58	15.48	-1.45	-64.89	25.13	-2.58
Answer Agreement * Syllable length * Content Predictability	-14.47	31.76	-0.46	22.05	55.15	0.40
Answer Agreement * Syllable length * Final Word Rate	15.52	19.32	0.80	40.77	42.04	0.97
Answer Agreement * Syllable length * Content Predictability *	-54.91	39.12	-1.40	-159.43	102.70	-1.55
Final Word Rate						
Answer Agreement * Question Duration	-49.81	21.54	-2.31	22.80	31.51	0.72
Answer Agreement * Question Duration * Content Predictability	-17.22	46.40	-0.37	61.68	55.24	1.12
Random effect	Yes responses			No responses		
		Variance	SD		Variance	SD
Item (Intercept)		8946	94.58		5737	75.74

Item (Final Word Rate)	0	0.00	0	0.00
Participant (Intercept)	51340	226.60	57116	238.99
Participant (Content Predictability)	6216	78.84	15740	125.46
Participant (Final Word Rate)	1224	34.99	5196	72.08
Participant (Syllable length)	0	0.00	897	29.95
Participant (Content Predictability * Final Word Rate)	0	0.00	0	0.00
Participant (Content Predictability * Syllable length)	0	0.00	4261	65.27
Participant (Final Word Rate * Syllable length)	0	0.00	0	0.00
Participant (Content Predictability * Final Word Rate * Syllable length)	0	0.00	0	0.00

---

Appendix B: Lists of stimuli used in the two experiments. Completions chosen from the pre-test are *italicized*

Table B1: Stimuli used in Experiment 1. All stimuli were unpredictable in content, and all final words consisted of a single syllable.

Stimulus
Do most people have two <i>jobs</i> ?
Are you happy when the weather is <i>dull</i> ?
Have you ever been bitten by a <i>cat</i> ?
Do you drink a lot of <i>juice</i> ?
Would you like to go for a walk in the <i>rain</i> ?
Do you like studying in the <i>dark</i> ?
Do cats have two <i>heads</i>
Do you find lectures very <i>dull</i> ?
During the summer, do you like spending time at the <i>house</i> ?
Have you ever visited the city of <i>Rome</i> ?
In your opinion, are you bad at <i>golf</i> ?

Are you allergic to *air*?

Do you have a sore *thumb*?

Have you ever been camping in the *rain*?

Do you like spending time at the *bar*?

Is chocolate your favorite *thing*?

Do kangaroos have two *ears*?

Do you have a good relationship with your *mum*?

Have you ever flown a *drone*?

Would you like to live in a different *home*?

Would you like to learn a new *phrase*?

Have you ever played a game of *cards*?

Do you have small *teeth*?

Do you need to go to the supermarket to buy some *wine*?

Do you sleep before *two*?

Do you have a big *heart*?

Do you have a pet *horse*?

Is an apple the same colour as a *rose*?

Do you often feel *stressed*?

Do you ever go to the *pub*?

At the weekend, did you do something nice for your *aunt*?

Do you have a lot of *cash*?

Do you often skip *meals*?

Would you like to go running in the *rain*?

Are there a lot of females in your *job*?

Do babies often cry when they are *young*?

Do you have four *phones*?

Have you seen my new *cat*?

Have you ever squashed a spider with a *map*?

Have you been sightseeing in *Skye*?

Is your handwriting *bad*?

Have you ever drawn a picture of a *whale*?

Have you ever had to apologise to your *boss*?

Is a pear the same colour as a *grape*?

Can you play a game of *cards*?

Do you have any *pets*?

Have you ever missed a *date*?

This morning, did you eat *eggs*?

Did you watch the tennis at *noon*?

Have you ever hurt yourself on a *plane*?

Would you like to get a new *ship*?

Do you need to buy some *shoes*?

Do you think you are good at *maths*?

Do you feel *cold*?

Have you ever failed an exam in *maths*?

Have you ever dyed your hair *blonde*?

Have you ever listened to music at a *rave*?

Would you like to take an evening *class*?

Do you ever worry about being *sick*?

Do you like to eat a lot of *sweets*?

Did you pay for your own *car*?

Should I buy a new suit for my *ball*?

For Christmas dinner, do you eat *ham*?

Have you ever been on a *date*?

Do you know how to cook *well*?

Would you like to get a *bird*?

Would you like another *car*?

Have you ever had an argument with your *dad*?

Do you have more than three *cats*?

Do you live far away from the *sea*?

Do you have high *heels*?

Can I give you a *book*?

Should I buy a nice new dress for my *ball*?

Do you spend a lot of money on *books*?

Tonight, can we stay out until *two*?



This morning, did you wake up at *noon*?

When travelling, do you get *lost*?

At University, do you study *maths*?

Can you buy me a *car*?

Do you often walk to *town*?

Do you watch a lot of *sport*?

Do you like my *car*?

Are you a big fan of *cheese*?

Do you think being a vegetarian is *cool*?

Are you free to go to the *beach*?

Would you like to have an afternoon *nap*?

Do you wear a *kilt*?

Is your favorite food *thai*?

Are your parents *nice*?

Have you ever had a bad *grade*?

In your opinion, do you think you are a good *friend*?

Are you shorter than your *dad*?

Is your hair very *fine*?

Is a grape different from a *plum*?

Have you ever won a game of *pool*?

Do you believe in *love*?

Would you like to see a picture of my *niece*?

In the past, have you had a lot of different *jobs*?

Do you have two *kids*?

Have you ever watched a game of *chess*?

Do you enjoy going to the *park*?

Have you ever seen a big *bird*?

Have you ever seen a wild *swan*?

Would you like to make an appointment with the *nurse*?

Do you own a *boat*?

Today, do you think I should wear a *tie*?

Tomorrow, would you like to wear a *watch*?

Would you like to go on holiday to *Greece*?

Are you very fond of *wine*?

When travelling, have you ever been on a *horse*?

Would you like to start attending classes on *time*?

Would you like to paint your *fence*?

Do you have a *dog*?

Do you have poor *health*?

Is your job *tough*?

Do tigers have big *heads*?

Would you like to travel to *Spain*?

Do you need a new *car*?

Do you want to buy a new *horse*?

For your age, are you *wise*?

Do you think you are *bad*?

Is an orange the same colour as a *peach*?

Do you think exercising is *fun*?

Have you had a long *trip*?

Table B2: Stimuli used in Experiment 2, broken down by content predictability and the syllable length of the final word.

Content Predictability	Syllable length	Stimulus
Predictable	1	Do chickens lay <i>eggs</i> ?
		Do dogs have four <i>legs</i> ?
		Is Paris the capital of <i>France</i> ?
		Does the president of America live in The White <i>House</i> ?
		Is 007 also known as James <i>Bond</i> ?
		Is the statue of liberty in New <i>York</i> ?
		Does the dentist tell you to brush your teeth twice a <i>day</i> ?
		Do you wash your hair every <i>day</i> ?
		Have you ever seen a spider with less than eight <i>legs</i> ?

Are pandas the colours black and *white*?

To pay for your studies, did you take out a *loan*?

Is a unicorn a horse with a *horn*?

Is a banana a *fruit*?

Is platform nine and three quarters at Kings *Cross*?

2

Is Harry Potter's best friend called Ron *Weasley*?

Is red your favorite *colour*?

Do genies grant *wishes*?

Did the titanic sink after hitting an *iceberg*?

Does the Queen live in Buckingham *Palace*?

Is Andy Murray a Scottish tennis *player*?

Do you think most students will pass their *exams*?

Will I need to buy a stamp before posting a *letter*?

Does the River Thames run through *London*?

Have you ever lived in a different *country*?

Is summer your favorite *season*?

Is a young cat called a *kitten*?

At University, are you a Psychology *student*?

3

Is a piano a musical *instrument*?

Is your favorite Jane Austen novel *Pride and Prejudice*?

Do you celebrate Christmas on the twenty fifth of  
*December*?

Is a trumpet a musical *instrument*?

Is Theresa May the prime *minister*?

Do you celebrate New Years eve on the thirty first of  
*December*?

Unpredictable

1

Are dogs your favorite *animal*?

Do you like studying in the *library*?

Do you often skip *lunch*?

Is your favorite food *fish*?

Do you have a sore *foot*?

Do most people have two *eyes*?

Have you been sightseeing in *France*?

Do you spent a lot of time on your *own*?

Do you spend a lot of money on *beer*?

Would you like to go running in the *rain*?

Do you have four *pets*?

Do you own a *house*?

Do you enjoy going to the *gym*?

Have you ever had to apologise to your *Dad*?

Do you have more than three *friends*?

Do you ever go to the *pub*?

2

Have you ever watched a game of *cricket*?

Have you ever injured your *finger*?

Is an apple the same colour as a *cherry*?

Have you ever played a game of *scrabble*?

Have you ever won a game of *poker*?

Do you have a good relationship with your *father*?

Are you allergic to *peanuts*?

Do you think you are good at *singing*?

Do you have two *siblings*?

Have you ever seen a wild *lion*?

Is an orange the same colour as a *carrot*?

Do you need a new *passport*?

Have you ever drawn a picture of a *person*?

3

At the weekend, did you do something nice for your  
*family*?



Do you know how to cook *spaghetti*?

Do you often see your *family*?

Have you ever visited the city of *Manchester*?

Do you want to buy a new *computer*?

Do you have a big *family*?

In your opinion, are you bad at *listening*?

Have you ever seen a big *elephant*?

---